A Quantitative Method for Measuring Library User Journal Needs: A Pilot Study Using CD Plus MEDLINE Usage Statistics

Kathel Dunn, M.S.L.S., Cheryl Chisnell, M.A., Suzanne Szak, B.A., Dean F. Sittig, Ph.D.

The Informatics Center, Eskind Biomedical Library
Vanderbilt University, Nashville, Tennessee

Objective: To develop a quantitative method for measuring library user journal needs based on an analysis of bibliographic search results. Design: Retrospective bibliometric comparison of citation selections generated by users in the library. Measures: Number of times each journal was identified by library users during multiple bibliographic search sessions. Results: Library users identified 4907 journal titles. The top 200 journal titles accounted for 55% of the library user journal needs. Of the 1380 unique titles identified, 652 were selected once. Conclusion: Our pilot study demonstrated that analysis of bibliographic search results can be used to identify library user journal needs. Such a method could also be used to estimate user requirements for online, full-text scientific journals.

INTRODUCTION

One of the hottest topics in corporate America today is customer service. Many organizations have hired expensive management consultants to help them identify their customers, determine their needs, and identify the requirements of those needs. Libraries are no exception. For hundreds of years, librarians have struggled with the decision of which journals they should have in their collection to best serve their customers. Now that the National Library of Medicine indexes over 3600 different journals, the decision is even more difficult. Only the largest and best endowed libraries can even consider having all these journals in their collection.

We hypothesize that it may be possible to assess the library customer's needs by watching what they do and learning from their actions. Our scenario for a given user's actions is this: a searcher sits at a computer and accesses a bibliographic database of a particular field. The searcher looks up a topic and selects the citations that are the most relevant to his/her work.

This study was undertaken to develop a quantitative method for measuring these library user needs based on a retrospective bibliometric analysis of their bibliographic search patterns.

BACKGROUND

A review of the literature reveals a number of attempts to establish a relationship between the user and the library collection. In constructing the library collection, for example, librarians develop criteria for the selection and deselection of materials for a particular collection [1,2]. Although a criteria list may include as many as ten items, the library user is generally only referred to in an oblique manner: as the "demand" [1] or as the "communal" [3]. Where there is mention of meeting user needs, there is little description of exactly how these interests and needs are determined. Certainly, traditional criteria cannot be abandoned and collections built solely on user wants, but more objective data are required to help in the selection of the materials.

Other researchers have used bibliometric methods for assessing a collection and in turn, establishing a relationship between what a user wants and the library collection. Two of the methods used include citation analysis and journal use studies. Citation analysis can either examine the work of one author, or set of authors, to determine publication patterns or it can determine the publications that have the highest number of cited articles, i.e., the most heavily used journals [4,5]. Citation analysis is only available for work that has been published and does not address the needs of the unpublished researcher or practicing clinician utilizing the library's collection in support of patient care.

In a similar manner, researchers will select a subject, for example, pediatrics, and then determine which journals produce the highest concentration of articles on the given topic [6,7,8,9]. While work of this nature has provided valuable insight into

0195-4210/94/$5.00

publishing patterns and established the concept of core journals in a field, it presupposes that users needs were completely met by the cited works.

The next step in these lines of study is to establish a closer link between the user and the collection in a quantitative, objective manner. With the advent of computers, researchers have been able to take advantage of new technology to assess collection development practices. CD-ROM and database management play a major part in moving collection development toward more quantitative, objective methods [10].

DESIGN/METHOD

Data Selection The Eskind Biomedical Library at Vanderbilt provides access to MEDLINE using CD Plus' Ovid software. Access is available to any workstation connected to the Medical Center network. This includes 400+ shared workstations scattered throughout the hospital and clinics as well as twenty-eight public workstations in the library. For the purpose of this pilot study, bibliographic search data were obtained from all CD Plus' MEDLINE searches conducted in the library on one day (March 17, 1994), using the last five years of the MEDLINE database (MEDL).

Data Manipulation A usage report for one day's bibliographic searches in MEDL was created. See example below. A "C" program was written to extract data from this report.

File Access:
medl

Citations browsed:	159
Citations printed:	1
Citations downloaded:	1254
Sets created:	9

Last search sets:

1 - Diastolic Dysfunction.tw.	303
2 - compliance/ and ventric$.tw.	82
3 - relaxation.tw. and heart/	323
4 - (relaxation and left ventr$.tw.	740
5 - 1 or 2 or 3 or 4	1476
6 - limit 5 to english language	1186
7 - diastol.$.tw.	11533
8 - 6 and 7	754
9 - limit 8 to review articles	116

A large, single search (*Search1*) comprised of all 870 separate search statements recorded in that file was constructed. *Search1* was executed in CD Plus' MEDLINE (1990-1994) to re-create the results obtained by the original searchers. In this way the search revealed not only the number of citations retrieved, but also the citations themselves. Since the intent of this pilot study was to quantify library user needs, it was necessary to limit the results of *Search1* by making some assumptions about which of the retrieved citations were actually wanted by the user. These assumptions were written as rules that were used repeatedly over all search statements to eliminate the user's intermediate results yet retain his/her final results.

For example, a user searching for the effects of vitamins during pregnancy might construct the following search:

Example Search in MEDL

	Search Word	No. of articles found
1	Vitamins/	797
2	Pregnancy/	49033
3	1 and 2	75

We infer from the above search that the user wanted the citations from search statement number 3 and not the citations retrieved from either statement 1 or 2. That is, the user was only interested in the terms vitamins and pregnancy together. Thus, the results of the first line (797 hits) for vitamins and the second line (49033 hits) for pregnancy are only part of the process in obtaining the final result of 75 hits. A rule reflecting the search above would be written, "Do not use results of any search statements that are used in a subsequent search statement." This rule was modified and other rules added to it as more searches, including more complex searches, were examined. The rules used to assess *Search1* are listed below:

Rules

1. If the results of any search generate 100 hits or more, then do not use these results. The default printing limit at our library is 100 citations.
2. If a search statement is used in a subsequent search statement, then do not use the earlier intermediate results. Searches that are used in earlier lines most likely represent only a portion of a thought process.

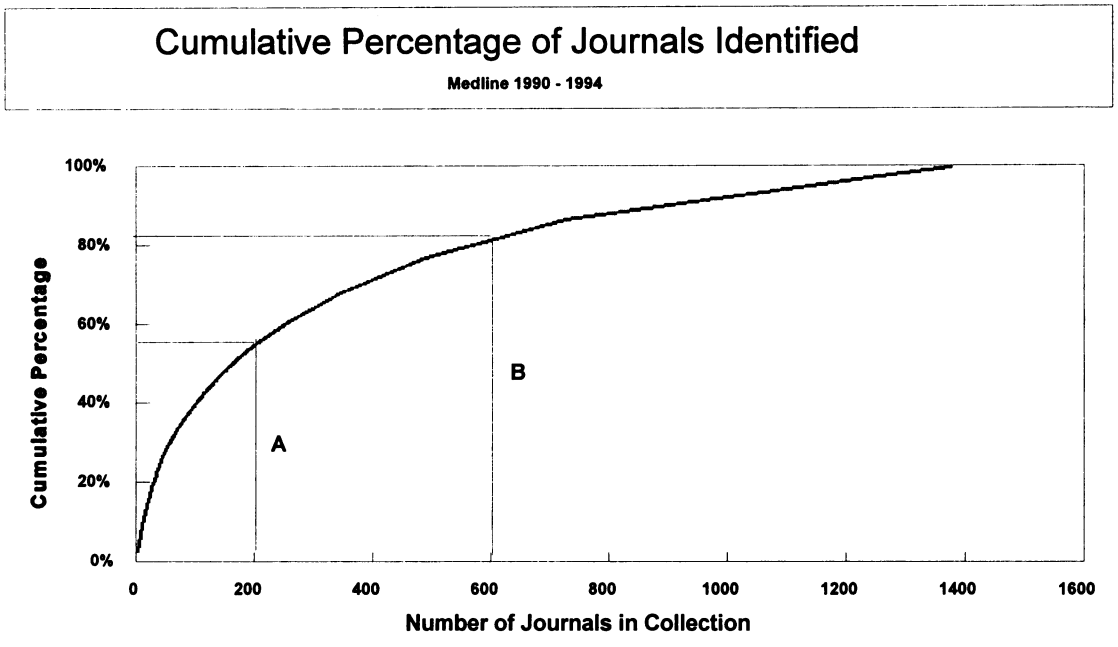

Figure 1

3. If a combined search statement results in 0 hits, then ignore Rule 2 and keep the results from the individual searches used in the combined search statement. The result of 0 hits probably required some rethinking on the part of the searcher, most probably requiring him/her to go back to the previously used statements.

Examples of the rules process:

	Search Word	No. of articles found
1	cocaine.tw.	3337
2	abruptio placentae/	103
3	1 and 2	10
4	from 3 keep 4,10	2

In the above search, lines 1 and 2 are eliminated because they have over 100 hits. Line 3 would normally have been kept as a result since it produced under 100 hits, but it is used again in line 4 -- "from 3 keep 4,10". Thus, line 3 is eliminated. The only results retained are those from line 4. Another example:

	Search Word	No. of articles found
1	(atropine and pa2).ti,ab,sh.	70
2	1 and muscarinic.ti,ab,sh.	48
3	from 2 keep 3,8,13,15,19-20,22-23, 25-26,28-29	12

In the above search all of the search statements produced hits of under 100, but line 1 was used in statement 2 and line 2 was used in statement 3, thereby eliminating the results of lines 1 and 2 from the final results set.

RESULTS

The Vanderbilt collection is split among several libraries, including the Stevenson Science Libray, the Heard Library and the Eskind Biomedical Library. Eskind Biomedical Library's journal collection numbers 2072. The MEDLINE database indexes 3600 journals and of that number, Vanderbilt holds 1884 titles. The total number of journal citations identified by library users was 4907. Of that number, 4140 are held in the Vanderbilt collection. A total of 1380 unique journal titles were identified by the users, of these 652 were selected once. Of the top 100 journals identified in this study, Vanderbilt holds 97% in its collection. The top 200 journal titles accounted for 55% of the library user journal needs.

Figure 1 shows the cumulative percentage of non-unique biomedical journals titles identified by library users from the current MEDLINE (1990-1994) file. Of particular interest is line A which shows the cumulative percentage of non-unique journal titles that came from the top 200 journals. Line B represents the top 600 journals.

Table 1. The top 32 journals identified.

Table 1

#	Journal Title	# of times ident
1	Journal of Biological Chemistry	133
2	Proc. Nat. Acad. of Sciences U.S.A	100
3	Development	42
4	Molecular & Cellular Biology.	41
5	Biochem.& Biophys. Res, Comm.	39
6	Journal of Bacteriology	39
7	Nature	34
8	Oncogene	34
9	Jour. of Pharm. & Exp.Ther.	33
10	Infection & Immunity	42
11	Nucleic Acids Research	33
12	EMBO Journal	32
13	Journal of Clinical Investigation	32
14	Science	38
15	FEBS Letters	32
16	Cancer Research	29
17	Biochemical Journal	28
18	Circulation	28
19	Radiology	27
20	American Journal of Physiology	27
21	European Journal of Pharmacology	26
22	Biochemistry	26
23	Endocrinology	25
24	Biochimica et Biophysica Acta	25
25	European Journal of Biochemistry	24
26	Journal of Virology	23
27	Magnetic Resonance in Medicine	23
28	Gastroenterology	23
29	Cell	23
30	New England Journal of Medicine	23
31	Ann. New York Acad. of Sciences	23
32	American Journal of Roentgenology	23

Figure 2 shows the number of times the library users selected unique journal titles.

DISCUSSION

The method developed in this pilot study represents a potential breakthrough in quantitatively defining the relationship between a library's users and its collection. By examining CD Plus' usage log of bibliographic search sessions, we can determine the final result sets identified by users as well as the contents of those sets. From a user's point of view, the final result set is exactly what s/he wants. We quantify what the user wanted and, more importantly, whether or not s/he obtained what s/he wanted. This quantitative method for measuring library user journal needs serves as an indicator for the development of a library's journal collection. Also, the method could serve as an indicator of whether or not a journal should be maintained in paper or received in an online, full text format.

Returning to the scenario described earlier, after the user has selected his/her chosen citations, s/he then attempts to retrieve them from whatever library collection is available. It is in stepping away from the computer that a user is frequently met with disappointment: the library collection does not have his/her selected items. The choice of selected items is usually modified and changed, based on the availability of items in a particular library's collection. At this point, knowledge of what a user wanted is forever lost.

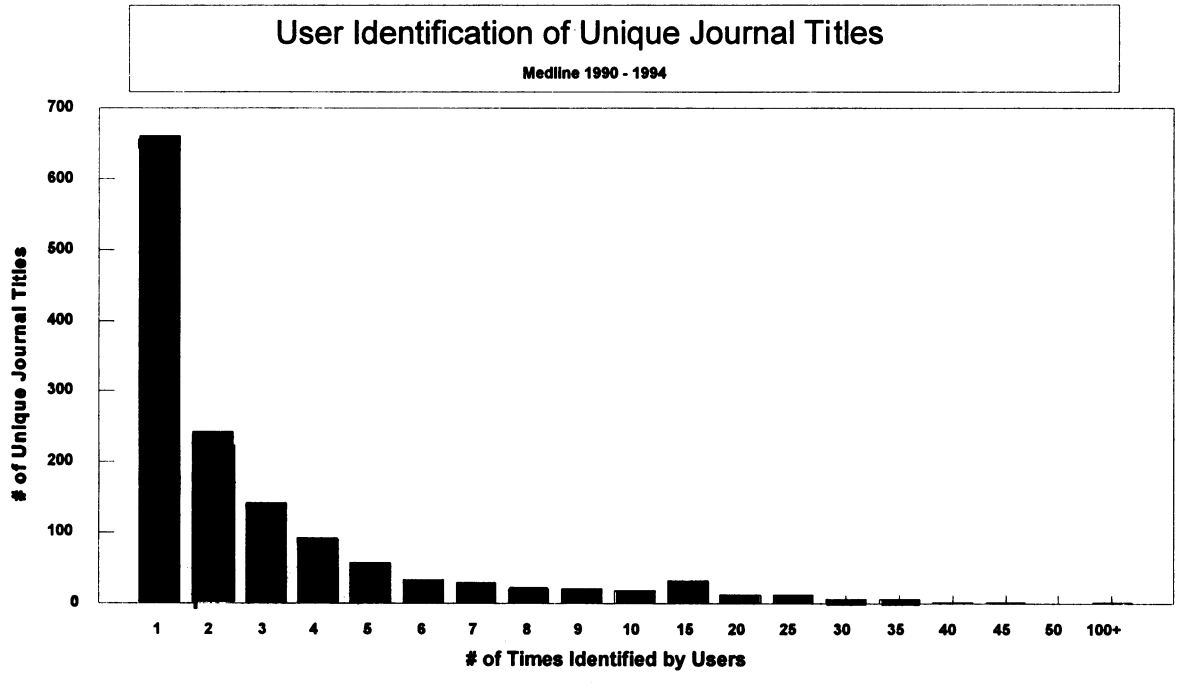

Figure 2

Limitations with this method

Data For the purposes of testing this method and assessing its feasibility, we used only one day's worth of MEDLINE (MEDL) searches conducted in the library. In order to draw more meaningful conclusions, a larger set of data is necessary. A longer time period with searches done by a wider variety of individuals will not only produce more data, but will further validate the use of this method as a collection development tool. Including other data in the study, such as journal usage studies and document delivery statistics, will further define and validate the relationship between the user and the collection.

Other research has concentrated on whether or not library users have chosen the most appropriate database and/or searched it well. Our method, designed to assist in the collection development process, concentrates on the customers' needs rather than on the process by which those needs are identified. Regardless of whether or not the searcher chose or searched well, under our rules, s/he still identified the particular journal titles that s/he wanted and hence that our library should have.

Rules The rules used in this method were crucial to the outcome. The rules determined what data were chosen and what were left out. Given that the rules were applied on search statements that represent the thought processes of any number of different searchers, there is inherent error. That is, not everything that a searcher wanted may have been included and, conversely, items that a searcher didn't want may have been included in the results. In creating the rules the decision was made to err on the side of inclusiveness. A larger data set would limit the effect of these errors on the final results.

System Features An important feature of CD Plus is its ability to "limit to local holdings." This feature has the potential to skew the data by 1) causing us to overestimate the percentage of non-unique journal titles held by Vanderbilt and 2) limiting our ability to measure accurately what library users actually want without regard to current library holdings.

To assess this potential problem, we counted the number of times that the "limit to local holdings" feature occurred in *Search1*. Of the 870 individual search statements, 30 contained this limit. If we assume that a complete search session contains, on average, 5 individual search statements, then only 17% of the search sets identified by users were affected.

Implications for further research

The next step in our research is to test this method with more data. We hope to also test this method against other collection development methods including citation analysis, journal use studies, inter-library loan and circulation statistics.

REFERENCES

[1]. J. Rutledge., L. Swindler. "The selection decision: defining criteria and establishing priorities." College and Research Libraries. 48(2):123-31, 1987.

[2]. RK Hunt. "Journal deselection in a biomedical research library: a mediated mathematical approach." Bulletin of the Medical Library Association. 78(1):45-8, 1990.

[3]. R. Atkinson. "The citation as intertext: toward a theory of the selection process." Library Resources and Technical Services. 28:109-19, April-June 1984.

[4]. ME Fang. "Journal rankings by citation analysis in health sciences librarianship." Bulletin of the Medical Library Association. 77(2):205-11, 1989.

[5]. L. Bronars., K. Branch. "Cost-cutting uses of new Scisearch feature." Database 13(6):53-8, 1990.

[6]. KS. Johnson., DJ Leising. "The literature of occupational therapy: a citation analysis study." American Journal of Occupational Therapy. 40(6):390-6, 1986.

[7]. RW Bohannon., DF Gibson. "Citation analysis of Physical Therapy: a special communication." Physical Therapy 66(4):540-1, 1986.

[8]. M. Berger., J. Devine. "Serials evaluation: an innovative approach." Special Libraries 81(3):183-8.

[9]. C. McDonough. "Measurement of the potential demand for academic and professional journals: a methodology." Journal of the American Society for Information Science 33(5):321-4, 1982.

[10]. J. Burnham., B. Shearer. J. Wall. "Combining new technologies for effective collection development: a bibliometric study using CD-ROM and a database management program." Bulletin of the Medical Library Association. 80(2):150-6.

This work was supported in part by grants R29 LM05284 and G08 LM05443 from the National Library of Medicine.